DOCUMENT RESUME

ED 237 576 TM 830 846

AUTHOR	Baker, Linda
TITLE	Children's Effective Use of Multiple Standards for Evaluating Their Comprehension.
INSTITUTION	Maryland Univ., Baltimore.
SPONS AGENCY	National Inst. of Education (ED), Washington, DC.
PUB DATE	21 Jul 83
GRANT	NIE-G-81-0100
NOTE	32p.
PUB TYPE	Reports - Research/Technical (143)
EDRS PRICE	MF01/PC02 Plus Postage.
DESCRIPTORS	Age Differences; Cognitive Processes; *Cognitive Style; *Comprehension; Elementary Education; *Evaluation Criteria; Prose; *Reading Strategies; *Self Evaluation (Individuals)
IDENTIFIERS	*Comprehension Monitoring; Embedding Transformations

ABSTRACT

Two experiments examined children's ability to apply three different standards for evaluating their understanding. Five-, seven-, nine-, and eleven-year-old children were presented with short narrative passages within which were embedded three types of problems (nonsense words, internal inconsistencies, and prior knowledge violations), each of which could only be identified if a specific standard of evaluation were used (lexical, internal consistency, and external consistency, respectively). Since the focus of the study was on the effectiveness with which children could apply the standards, rather than on the likelihood that they would spontaneously adopt and then apply them, the subjects were explicitly instructed in advance that their task was to find the "mistakes." Moreover, the subjects were given immediate feedback after each trial and a second opportunity to find any missed problems. Although older children used all three standards more effectively than younger children, overall problem identification was considerably better than that reported in non-instructed settings. The internal consistency standard was applied least effectively, but even the youngest children were able to use it. The results illustrate the need to consider comprehension monitoring skills with respect to specific standards of evaluation, rather than as a unitary phenomenon. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED237576

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
☒ This document has been reproduced as received from the person or organization originating it.
☐ Minor changes have been made to improve reproduction quality.
• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Children's Effective Use of Multiple Standards for Evaluating Their Comprehension

Linda Baker

University of Maryland Baltimore County

Running head: Multiple Standards

This research was supported in part by the National Institute of Education under grant number NIE-G-81-0100 and by a UMBC Summer Faculty Fellowship. I am grateful to Kris Kletkowicz and Kim Brady for their help in developing the materials, running subjects, and scoring the response protocols. I thank Bruce Lombardi for his careful comments on the manuscript. Requests for reprints should be sent to the author at the Department of Psychology, UMBC, Catsonville, MD 21228.

Date submitted: July 21, 1983

TM 830 846

Children's Effective Use of Multiple Standards

for Evaluating Their Comprehension

Date submitted: July 21, 1983

Abstract

Two experiments examined children's ability to apply three different standards for evaluating their understanding. Five-, seven-, nine-, and eleven-year-old children were presented with short narrative passages within which were embedded three types of problems (nonsense words, internal inconsistencies, and prior knowledge violations), each of which could only be identified if a specific standard of evaluation were used (lexical, internal consistency, and external consistency, respectively). Since the focus of the study was on the effectiveness with which children could apply the standards, rather than on the likelihood that they would spontaneously adopt and then apply them, the subjects were explicitly instructed in advance that their task was to find the "mistakes." Moreover, the subjects were given immediate feedback after each trial and a second opportunity to find any missed problems. Although older children used all three standards more effectively than younger children, overall problem identification was considerably better than that reported in non-instructed settings. The internal consistency standard was applied least effectively, but even the youngest children were able to use it. The results illustrate the need to consider comprehension monitoring skills with respect to specific standards of evaluation, rather than as a unitary phenomenon.

Children's Effective Use of Multiple Standards for Evaluating Their Comprehension

Because comprehension monitoring is an ability which has important educational implications, the degree to which children monitor their comprehension has become an issue of concern to many investigators. Educators have long argued that students should keep track of their understanding and should respond appropriately if they detect a failure to understand (e.g., Huey, 1906; Thorndike, 1917). However, more recently, empirical studies have demonstrated that children are surprisingly unlikely to carry out these evaluation and regulation activities (e.g., Garner, 1980; Markman, 1977, 1979; Paris & Myers, 1981). While the research has shown that older and better students monitor their comprehension more effectively than younger and poorer students, the latter group still shows considerable room for improvement.

The basic paradigm used in studies of comprehension monitoring is to introduce a problem or "error" of some sort into a prose passage and assess the subjects' ability to detect it. The rationale for this procedure is that subjects should notice the problems if they are keeping a careful check on their understanding. However, the fact that subjects typically are not informed in advance that the passages are problematic can lead to serious underestimation of their comprehension monitoring activities. This is because people tend to believe that the communications they receive will be true, complete, and informative (Grice, 1975) and consequently they attempt to make sense of any input, however confusing (cf. Baker, 1979). The few studies which have alerted subjects to the presence of problems generally report

higher levels of performance (e.g., Markman, 1979; Markman & Gorin, 1981).

The nature of the embedded problem may also have a strong influence on children's success or failure on the task. Researchers have used a number of different kinds of problems, such as nonsense words, violations of prior knowledge, and inconsistencies within the text itself. Detection of these different problem types requires the application of different criteria or standards of evaluation. In order to detect a nonsense word, a child must evaluate her understanding of individual word meanings. In order to detect a prior knowledge violation, a child must consider how the ideas in the text relate to what she already knows. And to detect an internal inconsistency, the child needs to evaluate the consistency of the ideas expressed in the passage. Thus, if a child fails to adopt a particular standard of evaluation, she will not notice the corresponding problems. This does not mean, however, that she did not evaluate her understanding along other dimensions. Since most studies have used only one type of embedded problem, they may present a misleading picture of children's comprehension monitoring skills.

There is some evidence that children are more likely to spontaneously adopt some standards than others. For example, Garner (1981) found that children were more likely to use a lexical standard than an internal consistency standard, as indexed by their superior identification of difficult vocabulary items and their lack of identification of any inconsistencies. And there is also evidence that children adjust their standards according to the instructions they receive. Markman and Gorin (1981) found that children who were set to evaluate for internal consistency identified more inconsistencies than falsehoods, while children set to evaluate for external consistency

identified more falsehoods than inconsistencies. However, we do not know whether children can keep several different standards in mind and use all of them effectively. Additionally, we have no information regarding developmental changes in the variety of standards children can use.

The present study was therefore designed to examine the effectiveness with which children of different ages can apply multiple standards for evaluating their understanding. Three specific standards were the focus of the investigation: lexical, internal consistency, and external consistency. Effective use of these standards was operationalized as success at finding embedded nonsense words, inconsistencies, and prior knowledge violations, respectively. The children were specifically informed that they would need to use these standards in order to find the "mistakes" in several passages. The instructions were made as explicit as possible because the concern was with children's ability to use the standards, not with whether they would spontaneously adopt them. Moreover, standard use was fostered during the experimental session by providing children with immediate feedback and a second opportunity to identify missed problems. Also of interest, in addition to problem identification, were the children's evaluative comments about other parts of the passages. These comments are important because they can reveal whether or not a particular standard was in fact present in a given child's repertoire, and if so, whether it was used selectively.

Two experiments were carried out. In the first, five-, seven-, and nine-year-old children listened to the passages which were read aloud by the experimenter. In the second, eleven-year-old children read the passages on their own and were later interviewed about their perceptions of standard

difficulty (e.g., Which type of problem was easiest to notice and why?). It was expected that the older children would apply all three types of standards more effectively than younger children, in line with previous findings. However, overall performance levels were expected to be higher than previously reported because of the specificity of the instructions and the provision of feedback wiith a second chance to find the problems. It was also expected that the lexical standard would be used most effectively and the internal consistency standard least effectively, an expectation based on the fact that these two standards differ considerably in the level of processing required for their application.

Experiment 1

Method

Subjects. A total of 53 5-, 7-, and 9-year-old children participated in the study. The two older groups of children were enrolled in the first and third grades of a suburban public school. There were 16 seven year olds ($\bar{X}$ age = 6 years, 11 months; nine boys) and 14 nine year olds ($\bar{X}$ age = 9 years, 0 months; seven boys). The youngest group was drawn from two different suburban preschools. There were 11 children from one preschool ($\bar{X}$ age = 4 years, 8 months; five boys) and 12 children from the second preschool ($\bar{X}$ age = 5 years, 1 month; 6 boys). All children were tested during the last two weeks of the school year except for the children in the first preschool, who were seen in January.[1]

Materials. The materials consisted of six short narrative passages dealing with topics and situations familiar to young children. Each passage was 7 to 9 sentences in length ($\bar{X}$ = 8.17) and contained from 69 to 82 words ($\bar{X}$

= 75). Each passage was modified to contain two of three different types of problems which could be detected by use of the appropriate standard of evaluation: nonsense words, prior knowledge violations, and internal inconsistencies. The nonsense words were selected from a list of two-syllable "paralogs" (Noble, 1952) and followed standard rules of English orthography. The nonsense words always replaced nouns that occupied the final position in a sentence. An example of a target sentence containing a nonsense word is, "Mrs. Johnson cooked the pancakes in a bladmer." The prior knowledge violations were created by introducing information that conflicted with world knowledge that children were assumed to possess. An example sentence is, "Jack always used a baseball bat to chop the wood." The internal inconsistencies involved two target sentences each, which were separated in the passage by one intervening sentence. The information contained in the two target sentences was contradictory. An example is, "He [a rabbit] had dark brown fur that was as soft as could be. He was very fluffy and had a beautiful tail. All the other rabbits wished they had his snow white fur." Two examples of representative passages appear in Table 1.

The problem types were systematically distributed throughout the passages such that each problem type appeared with every other problem type twice, once in first position and once in second position. The placement of the problems within the passages was restricted in two ways: No problems appeared in either the first or last sentence of any passage and when one of the problems was an internal inconsistency, the second problem was never placed in the intervening sentence.

Insert Table 1 about here

The experimental materials were developed from a larger set of 12 passages that were presented to 28 undergraduates drawn from an introductory psychology subject pool. In order to verify that the problems would be perceived as such by adults, the students were instructed to read the passages carefully in search of the three different types of problems. They were to underline the problems when they found them and write alongside the target sentence the type of problem they considered it to be. The response sheets were scored for the percentage of students correctly identifying each problem. The detection rate was consistently high, ranging from 82% to 100%, with an overall mean of 94%. The difference in detection rates among the three types of problems was nonsignificant (96%, 95%, and 93% for nonsense words, prior knowledge violations, and inconsistencies, respectively). A criterion of 93% detection was established as the cut-off for inclusion of the problem in the materials for the children. (In other words, no more than two of the 28 subjects could miss the problem.) A second criterion was that no more than two students could comment that a nonproblematic sentence in fact had some sort of problem. Taking these criteria into account, six of the passages were selected and modified for use as experimental materials.

The passages were arranged such that each type of problem appeared sequentially either within or between stories. The order of presentation of the passages was counterbalanced such that the first three passages for half of the subjects were the last three passages for the other subjects, and vice versa.

Procedure. The children were seen individually during school hours in a quiet room at their school. They were told that they would be listening to some stories that had mistakes in them and that these mistakes were there because the writers had not been careful enough when they wrote the stories. The children's job was to find the mistakes and tell the experimenter. The three types of problems or "mistakes" were described as follows: words that aren't really words; things that don't really happen the way it says in the story; and things in one part of the story that don't go with things in another part. Then a sample story containing one of each problem type was read and the children were asked to try to find the problems. Regardless of their performance, the experimenter then cited the embedded problems as concrete examples of the three different types that should be sought. The children were not told how many problems would be present in any given passage.

Since the primary purpose of the study was to assess children's evaluation skills under optimal circumstances, the task was structured to provide an opportunity for improvement in the child's performance within the session. To this end, the children were given two chances to find the problems. Each story was read aloud once by the experimenter and the children were asked to report whatever problems they noticed. They were encouraged to report anything that "didn't seem quite right" even if they were not sure it was a mistake. If the children did not report both problems that were present, they were told they had missed something and were asked to listen to the story again. To alleviate possible memory difficulties, they were encouraged to interrupt as soon as they noticed a problem. After the second reading, if the

children still did not report the problems, the experimenter identified them. This explicit feedback in essence served as additional instruction in using the standards of evaluation.

A written record was kept of the children's ongoing responses and the sessions were also tape recorded for later transcription. The sessions ranged in length from 10 to 25 minutes.

Scoring. The response protocols were scored for two different dependent measures: problem identification and standard application. Each problem was scored as correctly identified or not. If the children were not specific enough in their initial responses to permit scoring, the experimenter requested clarification. For example, if a child simply said the mistake was "the part about snow white fur," the experimenter asked, "What about snow white fur was wrong?" If the child then explained something to the effect that the story earlier said the rabbit had brown fur, the problem was scored as identified. If the child indicated that some other component was problematic (e.g., "There's no such thing as snow white fur"), the problem was scored as not identified. (Such comments were classified, however, as to the type of standard they revealed, as described below.) Responses were scored immediately, since the next step in the procedure was dependent on the judgment (e.g., if a child was correct on Trial 1, Trial 2 would not be necessary). However, all decisions were also checked after the tapes had been transcribed and an independent judge validated the decisions with 98% agreement.

All scoring for the standard application measure was completed after the tapes were transcribed. Each response the child made to any part of the story

was classified as to whether it revealed the use of a particular standard of evaluation. The classification scheme consisted of the three target standards, lexical, external consistency and internal consistency, as well as four others: propositional cohesiveness, structural cohesiveness, informational completeness, and syntax (cf. Baker, in press). It was also necessary to establish a category for "other" responses, because many of the five year olds made nonevaluative comments (e.g., "We have a dog too"). The protocols were scored by two independent judges, who resolved disagreements through discussion. Since children rarely used any of the nontargeted standards, the focus here will be on the target standards only.[2]

Results and Discussion

This part of the paper will be divided into two sections. The first section will focus on children's identification of the embedded problems and the second will examine the standards children used when identifying any parts of the passages as problematic. In a sense, the first section deals with the extent to which children's use of the three standards is similar to adult usage and the second section examines standard use from the individual child's perspective of what is or is not problematic.

Problem identification. Separate analyses were carried out for the number of problems the children identified after hearing the passage only once and for the total number of problems identified. In addition, a third analysis included trial as a factor in order to examine differential improvement when given a second opportunity to find the problemms. In this latter analysis, if a problem was identified on Trial 1, it was also scored as identified on Trial 2, even though the child had not actually been asked to identify it.[3]

Table 2 presents the mean number of problems of each type that the children identified after the first listening only (Trial 1) and after a second listening as necessary (Trial 2). Two 3 (age) x 3 (problem type) mixed analyses of variance were carried out using these two data sets, with age as a between-subjects factor and problem type as a within-subjects factor.

Insert Table 2 about here

Consider first the results for the first trial. Reliable main effects of age and problem type were obtained, and the interaction of the two factors was also reliable. As expected, older children were more successful at identifying all types of problems than younger children, $F(2,50) = 53.70$, $p < .001$. The 9 year olds identified an average of 3.26 problems of each type, the 7 year olds 2.16 and the 5 year olds 1.19. All comparisons between means were significant (Fisher's lsd = .39, $p < .05$). Also as expected, internal inconsistencies were least likely to be identified ($F(2,100) = 10.92$, $p < .001$. Subjects identified an average of 1.66 inconsistencies, 2.06 nonsense words, and 2.38 prior knowledge violations. All of the differences between means were reliable (Fisher's lsd = .28, $p < .05$). The interpretation of the effect of problem type is mediated by the interaction with age, $F(4,100) = 2.58$, $p < .05$. Whereas the 7- and 9-year-old children had comparable identification of nonsense words and prior knowledge violations, with poorer identification of inconsistencies, the 5 year olds had comparable identification of nonsense words and inconsistencies, with superior identification of prior knowledge violations (Fisher's lsd = .50, $p < .05$).

The Trial 1 data indicate that the expected developmental differences in problem identification were not constant across problem types. However, these pattern differences dropped out by the second trial, as revealed in the second analysis of variance. Again, the main effects of age and problem type were reliable, but the interaction was not ($F(2,50) = 36.26$, $p < .001$; $F(2,100) = 5.54$, $p < .01$; $F(4,100) = 1.43$, $p > .10$, respectively). Nine year olds identified more problems of each type (3.91) than seven year olds (3.29), who in turn identified more than five year olds (2.19). Internal inconsistencies were still least likely to be identified (2.66), and identification of nonsense words and prior knowledge violations did not differ (3.15 and 3.11, respectively).

The data included in Table 2 clearly indicate improvements in problem identification when children were given feedback and a second opportunity to listen to and respond to the passages. To test for differential improvement as a function of age or problem type, a third analysis of variance was carried out which included trial as a factor. The main effect of trial and the interaction with age were reliable, $F(1,50) = 165.74$, $p < .001$ and $F(2,50) = 4.23$, $p < .05$, respectively. All subjects identified more problems when they were given a second attempt to find them. The 5 and 7 year olds showed greater gains than the 9 year olds, whose initial levels of performance were so high that there was less room for improvement. Neither the interaction of trial with problem type nor the triple interaction were reliable. These results demonstrate that even the youngest children were able to re-evaluate the passages and detect problems in material that had previously seemed non-problematic.

Standard applications. The dependent measure in the above analyses was the number of times a child identified the intended problems. The analysis provides no information as to whether the children perceived other segments of the passages as problematic. The fact that these segments were not intended to contain problems is irrelevant; what is important is how the children interpreted them and their subsequent evaluation of that interpretation. In other words, the focus on problem detection per se may underestimate children's ability or propensity to apply the different standards of evaluation. In this section, we will consider evidence of standard use throughout the entire testing session.

Table 3 shows the mean number of times children applied each of the three targeted standards. The figures reflect standard use both in the service of problem detection and in the evaluation of nontarget information. An analysis of variance was carried out, with the depencent variable the total number of times each child applied each standard. Reliable main effects of age and type of standard were revealed, as was an age by standard interaction. Overall, the 5-year-old children had fewer standard applications than the 7- and 9-year-old children who did not differ, $F(2,50) = 20.69$, $p < .001$. The internal consistency standard was applied less frequently than the lexical standard which in turn was applied less frequently than the external consistency standard, $F(2,100) = 10.08$, $p < .001$ (Fisher's lsd = .44). Interpretation of the main effects is qualified by the interaction, $F(4,100) = 2.84$, $p < .05$. Although the 7 year olds did not differ from the 9 year olds in overall standard use, they used the internal consistency standard less often and the external consistency standard more often. They were actual-

comparable to the 9 year olds only in their frequency of use of the lexical standard. Also contributing to the interaction was the fact that the 9 year olds used all three standards equally often.

Comparison of the data in Table 3 with the Trial 2 data in Table 2 provides information about the effectiveness with which children used the different standards from an adult perspective. Recall that adults consistently identified the target problems and rarely identified other aspects of the text as problematic. The 9 year olds behaved much as the adults did, with near perfect problem identification and few additional problems reported. The younger children reported fewer problems, but the difference between the number of times a standard was applied and the number of corresponding problems identified was of relatively small magnitude. This indicates that the children were not simply guessing and identifying anything as problematic simply to comply with task demands; rather, they were applying the standards selectively. The largest discrepancy between number of problems identified and number of times a standard was applied was shown by the 7 year olds for prior knowledge violations and the external consistency standard. On average, each child challenged the truth of 1.63 facts in addition to the targeted facts. But considering that there were a total of 49 sentences, each of which contained at least one challengeable proposition, this still reflects selectivity in standard use. Nevertheless, the 7 year olds did have a greater propensity to challenge the external consistency of passage statements than either the younger or older children.

Additional insight into children's use of the different standards can be gained by examining patterns of individual use. Are there certain standards

that are less likely than other standards to be used at all? Are younger children less likely to apply any particular standards than older children? To answer these questions, each child's transcribed protocol was examined for presence or absence of each standard. This procedure revealed that all 7- and 9-year-old children used every standard at least once and that most of the 5 year olds did too. All of the 5 year olds used the external consistency standard, all but one used the lexical standard, and all but three used the internal consistency standard. Clearly, then, virtually all children used the three different standards. Moreover, if a child used a standard at all he or she also usually detected at least one of the corresponding problems. In only two instances did a child use a standard but fail to report any problems (both children were 5 year olds: once the standard was the lexical, the other time the standard was the internal consistency). In sum, it seems safe to conclude that all three standards were present in the repertoires of children of all ages and that these standards could be used effectively and selectively.

Experiment 2

Experiment 2 was designed as a replication and extension of Experiment 1 using an older group of subjects. The 11-year-old subjects in this experiment were asked to read the passages on their own rather than listen to them, but in all other respects the task was the same. The change from listening to reading, which was not expected to affect performance significantly, was made to permit a better comparison with studies reporting poor problem detection among 10- and 11-year-old children who read passages without being told in advance that problems were present (e.g., Garner, 1980; Paris & Myers, 1981). After the subjects finished reading and responding to all six passages, they

were questioned as to whether they thought some problem types were easier or harder to detect and were asked to explain their answers. The explanations were of particular interest for what they might reveal about children's metacognitive knowledge about comprehension monitoring.

Method

Subjects. The subjects were 18 children (11 girls) enrolled in the fifth grade at the same suburban elementary school as the participants in Experiment 1. The mean age of the subjects was 11 years, 0 months.

Materials. The materials were the same as those used in Experiment 1. Each passage was typed double-spaced on a 12.7 x 20.3 cm note card with the title at the top.

Procedure. The procedure was similar to that of Experiment 1. The primary difference was that the subjects read the passages on their own instead of listening to them. To make the task more comparable to the listening task, subjects covered up each line of text with a blank note card after they read it. When they finished the entire passage, they placed the card down on top of the passage and reported whatever problems they had noticed. They were given immediate feedback and re-read the passage if they missed any problems. They were encouraged to report the problems as soon as they found them on this second trial. If they still failed to report a problem, the experimenter identified it for them.

After the subjects finished reading all passages, they were asked whether they thought some problem types were easier/harder to notice than others. Half of the subjects were asked the question with the word easier, half with the word harder. If they gave an affirmative answer, they were asked to

explain it. They were then asked the alternative form of the question (harder/easier) and were again asked to explain.

All sessions were taped and later transcribed. The scoring of the responses and the transcribed protocols was carried out in the same manner as in Experiment 1.[4]

Results and Discussion

The first section in this part of the paper will focus on problem identification. The second section will focus on children's reports of perceived problem difficulty. A final section will consider standards used in the evaluation of non-target as well as target problems.

Problem identification. A single analysis of variance was carried out on the number of problems identified, with trial and problem type as within-subjects factors. The relevant data are included in the last two rows of Table 2. The main effect of problem type was reliable, $F(2,30) = 13.60$, $p < .001$, with fewer inconsistencies identified than either nonsense words or prior knowledge violations (Fisher's lsd = .46, $p < .05$). The main effect of trial was reliable, $F(1,15) = 35.31$, $p < .001$, illustrating that the 11-year-old children also benefitted from a second opportunity to evaluate the passages. The extent of the improvement varied according to problem type, as evidenced in the trial by problem type interaction, $F(2,30) = 10.88$, $p < .001$. On trial 1, nonsense words were better identified than prior knowledge violations, which in turn were better identified than inconsistencies (Fisher's lsd = .46, $p < .05$). On Trial 2, the three problem types were equally well detected, with the gains from Trial 1 to Trial 2 reliable for the inconsistencies and prior knowledge violations. The lack of a significant

difference for the nonsense words reflects the fact that subjects were near ceiling on Trial 1, leaving little room for improvement on Trial 2. There were no other reliable main effects or interactions.

Perceived problem difficulty. After completion of the task, the children were asked which, if any, problem types were easier to detect and which were harder to detect. Of the 18 subjects, all but two indicated that the nonsense words were the easiest. The labels they used varied, from "nonsense words" to "weird words" to "misspelled words" to "words that aren't honest to gosh words that you find every day," but it was always clear what was meant. The remaining two children indicated specific problems in their responses. One mentioned "baseball bats" as easiest (a prior knowledge violation), and the other said that "most were easier to find than the one about checkers" (an internal inconsistency).

In response to the "harder" question, 11 children reported that the inconsistencies were hardest. Usually this was indicated by giving a specific example (e.g., "like when they said soft brown fur and then they said white"). Only three children explained what it was about the inconsistencies that made them harder. Two of these children attributed the difficulty to memory factors; e.g., "You need to keep one word in your head and then remember the other; you need to try to remember both." The third identified a particular reading strategy as the source of her difficulty: "Sometimes when I read I skip over words, like big-small, so I didn't pick them up." Of the seven children who did not report inconsistencies as the most difficult, two said prior knowledge violations were the hardest and five said "nothing."

In sum, the reports of perceived difficulty parallel the actual difficulty. However, the reports were probably influenced by prior performance, since the children had received feedback. In fact, usually the children cited as most difficult one of the specific problems they had missed. It would be informative to ask for children's judgments before they carry out such an evaluation task.

Standard applications. Table 3 includes the mean number of times each child applied the three standards of evaluation regardless of whether the response was to an intended problem. A one-way analysis of variance with standard type as a within-subjects factor revealed no difference in the frequency with which the different standards were used, $F(2,34) < 1$. This parallels the lack of a difference in the identification of the three types of problems by the second trial. The similarity in the means for problem identification and standard application indicates that the 11 year olds used the standards very selectively, in a manner comparable to the adults. They never identified any words as problematic that were not intentionally so and only twice did subjects identify perceived violations of prior knowledge that were not deliberately introduced. Additionally, the internal consistency standard was applied to nontarget information on only five occasions. Moreover, as would be expected given the patterns of standard use among the older children in Experiment 1, there were no subjects who failed to use any of the target standards, nor were there any who failed to report any type of problem.

General Discussion

The present experiments have shown that, across a wide age range, children have considerable skill at using several different standards to evaluate their

understanding. Although older children were more successful than younger children, the absolute levels of performance were much higher than has typically been reported in the literature. There are several important reasons for this difference: (1) The children were explicitly informed that problems would be present and they were given several examples of each type. Failure to provide this information in advance can lead to serious underestimation of comprehension monitoring skills (Baker, 1979; Stein & Trabasso, 1982; Winograd & Johnston, 1981). (2) The children were given immediate feedback after each attempt to find a problem. This feedback in essence served as instruction in evaluation. (3) The children were asked to re-evaluate the passages for missed problems, giving them an opportunity to improve their performance. (4) Three different types of problems were used, in recognition of the fact that comprehension monitoring requires the use of multiple standards, some of which may be easier to apply than others.

Of the three standards selected for study, the internal consistency standard was expected to be most difficult to apply and the data supported this expectation. Children ranging in age from 5 to 11 were less likely to identify internal inconsistencies in the passages than they were the other problem types. Evaluating text for internal consistency is a cognitively demanding task because it requires that the reader or listener first integrate the relevant text propositions. If the propositions are not adjacent, they may not be simultaneously active in working memory and so a long term memory representation must be accessed (Kintsch & van Dijk, 1978). The likelihood that adults will carry out this integration during reading decreases with increasing distance between the propositions (Walker & Meyer, 1980).

Moreover, Johnson and Smith (1981) have shown that younger children seem not to integrate widely separated propositions even when the necessary premises are available in memory. In view of this evidence, it is hardly surprising that the internal consistency standard is less frequently applied.

It was also expected that the lexical standard would be easiest to apply because it can operate on individual words; it is not necessary to consider the relationships between ideas. Although Paris and Meyers (1981) reported poor detection of nonsense words by fourth graders, their finding may have reflected an unwillingness on the part of the subjects to admit ignorance of seemingly simple words rather than ineffective use of a lexical standard. The present data support this alternative interpretation; both the 9 and 11 year olds had near perfect nonsense word detection, and the 7 year olds were also very successful. The one surprise was that the 5-year-old children initially identified no more nonsense words than internal inconsistencies. However, when given feedback that they had missed something on the first trial, their identification rate increased from 25% to 58%.

No specific predictions were made regarding the external consistency standard, though there is reason to suspect it would be of intermediate difficulty to use. Since evaluation of external consistency requires consideration of how ideas in the text relate to what one already knows, greater cognitive effort is probably required to detect a prior knowledge violation than a nonsense word. And, on the assumption that the relevant prior knowledge is probably more accessible in long term memory than previously encountered text propositions, prior knowledge violations should be more detectable than internal inconsistencies. The data do not provide clear

evidence regarding these possibilities. Internal inconsistencies were in fact less frequently reported than prior knowledge violations by all children. But the 5 year olds were initially better at identifying prior knowledge violations than nonsense words, while the 11 year olds were initially better at identifying nonsense words than prior knowledge violations. By the second trial, however, there were no differences for any children in detection of nonsense words and prior knowledge violations. Interpretation of this finding is difficult because overall detection levels were so high. If the detection rates had been lower, perhaps there would have been "room" for differences to manifest themselves. Increasing the difficulty of the materials could have such an effect, but it might obscure differences among the younger children. The risk of encountering either ceiling or floor effects is a familiar one to researchers who wish to study developmental change across a wide age range. However, since the primary goal in this study was to show what children can do under optimal circumstances, ceiling effects were regarded as a necessary evil.

Finally, the results extend those of Markman and Gorin (1981) in showing that not only can children adjust their standards of evaluation, they can keep several in mind simultaneously. There was, however, a very striking difference in the performance levels of the children in the two studies, despite the apparent similarity of materials and task demands. Even when specifically instructed to look for a particular type of problem, the 8- and 10-year-old children in Markman and Gorin's study had much lower detection rates than the children of comparable ages in the present study. The difference may well lie in the feedback that was provided after each attempt

to find a problem; the opportunity to "try again" led to significant gains. Overall, the results are encouraging in suggesting that children can be induced to monitor their comprehension more effectively with minimal intervention.

References

Baker, L. Comprehension monitoring: Identifying and coping with text confusions. Journal of Reading Behavior, 1979, 11, 363-374.

Baker, L. How do we know when we don't understand? Standards for evaluating text comprehension. In D. L. Forrest, G. E. MacKinnon, & T. G. Waller (Eds.), Metacognition, cognition, and human performance. New York: Academic Press, in press.

Flavell, J. H., Speer, J. R., Green, F. L., & August, D. L. The development of comprehension monitoring and knowledge about communication. Monographs of the Society for Research in Child Development, 1981, 46, No. 5 (Whole No. 192).

Garner, R. Monitoring of understanding: An investigation of good and poor readers' awareness of induced miscomprehension of text. Journal of Reading Behavior, 1980, 12, 55-64.

Garner, R. Monitoring of passage inconsistency among poor comprehenders: A preliminary test of the "Piecemeal Processing" explanation. Journal of Educational Research, 1981, 74, 159-162.

Grice, H. P. Logic and conversation. In P. Cole & J. L. Morgan (Eds.), Syntax and semantics (Vol. 7): Speech acts. New York: Academic Press, 1975.

Harris, P. L., Kruithof, A., Terwogt, M., & Visser, T. Children's detection and awareness of textual anomaly. Journal of Experimental Child Psychology, 1981, 31, 212-230.

Huey, E. B. The psychology and pedagogy of reading. Cambridge, Mass.: M.I.T. Press, 1968. (Originally published, 1908)

Johnson, H., & Smith, L. B. Children's inferential abilities in the context of reading to understand. Child Development, 1981, 52, 1216-1223.

Kintsch, W., & van Dijk, T. A. Toward a model of text comprehension and production. Psychological Review, 1978, 85, 363-394.

Markman, E. M. Realizing that you don't understand: A preliminary investigation. Child Development, 1977, 46, 986-992.

Markman, E. M. Realizing that you don't understand: Elementary school children's awareness of inconsistencies. Child Development, 1979, 50, 643-655.

Markman, E. M., & Gorin, L. Children's ability to adjust their standards for evaluating comprehension. Journal of Educational Psychology, 1981, 73, 320-325.

Noble, C. E. An analysis of meaning. Psychological Review, 1952, 59, 421.

Paris, S. G., & Meyers, M. Comprehension monitoring, memory, and study strategies of good and poor readers. Journal of Reading Behavior, 1981, 13, 5-22.

Stein, N. L., & Trabasso, T. What's in a story: An approach to comprehension and instruction. In R. Glaser (Ed.), Advances in instructional psychology, Vol. 2. Hillsdale, N.J.: Erlbaum, 1982.

Thorndike, E. L. Reading as reasoning: A study of mistakes in paragraph reading. Journal of Educational Psychology, 1917, 8, 323-332.

Walker, C. H., & Meyer, B. J. F. Integrating different types of information in text. Journal of Verbal Learning and Verbal Behavior, 1980, 19, 263-275.

Winograd, P., & Johnston, P. Comprehension monitoring and the error detection paradigm. Journal of Reading Behavior, 1981, 14, 61-74.

Footnotes

[1]Four additional five year olds served as subjects but they were dropped from the study because their data records were incomplete due to a tape-recorder malfunction. Information regarding the number of problems identified was available from the written records made at the time of testing, but information about all other instances of standard use depended on transcriptions of the tapes, which were unavailable.

[2]One 5 year old made a comment reflecting evaluation of informational completeness and two 7 year olds used the propositional cohesiveness standard. In addition, 14 5 year olds made one or more nonevaluative comments about the passages.

[3]Prior to carrying out a full analysis of the data, the comparability of the two different samples of preschoolers was assessed. An analysis of variance was carried out on the data provided by these children alone, with school as a between-subjects factor. There was no main effect of school ($F(1,20) = 1.19$) nor were there any interactions of school with the factors of interest. Accordingly, the data from the two samples were combined for subsequent analysis. A second preliminary analysis, using the full data set, included gender as a factor. The main effect was not reliable nor were any interactions with gender. The data were therefore collapsed over gender for ease of exposition.

[4]Use of non-target standards was infrequent among 11 year olds as well. Two students commented on the structural cohesiveness of the passages and one on the informational completeness.

Table 1

Examples of Passages with Embedded Problems

Albert the Pretty Rabbit

Once there was a rabbit named Albert. He had dark brown fur that was as soft as could be.[a1] He was very fluffy and had a beautiful tail. All the other rabbits wished they had his snow white fur.[a2] Albert liked to eat in Farmer Smith's garden. Lots of good things grew in the garden. But Albert especially liked the ice cream that grew there.[b] Farmer Smith did not like rabbits to eat his food. Albert was lucky he never got caught.

Jack's Life

There lived a man who had three sons. The youngest son was named Jack. Jack was tall and had brown curly welkins.[c] Every morning Jack chopped wood for his family. He always used a baseball bat to chop the wood.[b] He had to do it quickly on school days so he wouldn't be late. Jack liked school very much. He wanted to be a teacher when he grew up.

a1, a2 - sentences contain an internal inconsistency

b - sentence contains a prior knowledge violation

c - sentence contains a nonsense word

Table 2

Mean Number of Problems Identified

			Type of Problem		
	Age	Trial	Nonsense Word	Prior Knowledge Violation	Internal Inconsistency
Experiment 1					
	5	1	1.00	1.61	0.96
		2	2.30	2.48	1.78
	7	1	2.38	2.62	1.50
		2	3.63	3.31	2.94
	9	1	3.43	3.36	3.00
		2	4.00	3.93	3.79
Experiment 2					
	11	1	3.77	3.28	2.55
		2	4.00	4.00	3.77

Table 2

Mean Number of Times Children

Applied Each Standard

		Type of Standard		
	Age	Lexical	External Consistency	Internal Consistency
Experiment 1				
	5	2.78	3.57	2.17
	7	4.38	4.94	3.25
	9	4.07	4.14	4.21
Experiment 2				
	11	4.06	4.00	4.11